DOCUMENT RESUME

ED 139 832 TM 006 325

AUTHOR Abrami, Philip C.; And Others
TITLE The Generalizability of Student Ratings of Instruction.
PUB DATE [Apr 77]
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; Achievement Tests; Correlation; *Generalization; Higher Education; Individual Differences; Models; *Rating Scales; *Research Design; *Student Characteristics; *Student Evaluation of Teacher Performance; *Teacher Characteristics; Test Results; Test Validity

ABSTRACT

Studies of teacher rating form (TRF) generalizability have focused primarily on two classes of variables: student characteristics and teacher characteristics. Since the goal of most TRF researchers is to show that TRF scores correlate with teacher-produced student achievement, a suitable generalizability study would investigate the relationship between student or teacher characteristics and the correlation between TRF scores and achievement scores. A TRF can only be considered generalizable if the validity coefficient is unaffected by student or teacher characteristics. Since researchers have not studied the generalizability of the TRF validity coefficient, the generalizability of TRFs is still an open question. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED139832

Session # 20.08

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Generalizability of Student Ratings of Instruction

Philip C. Abrami, Les Leventhal, and Raymond P. Perry

University of Manitoba

Running Head: Generalizability

Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977.

Requests for reprints should be addressed to:
Philip C. Abrami, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

TM006 325

Abstract

Studies of teacher rating form (TRF) generalizability may focus on three classes of variables: student characteristics, teacher characteristics, or situational characteristics. Since the goal of most TRF researchers is to show that TRF scores correlate with teacher-produced student learning, a suitable generalizability study would investigate the relationship between student, teacher, or situational characteristics and the correlation between TRF scores and learning scores. A TRF can only be considered generalizable if the validity coefficient is unaffected by student, teacher, or situational characteristics. Since researchers have not studied the generalizability of the TRF validity coefficient, the generalizability of TRFs is still an open question.

The Generalizability of Student Ratings of Instruction

Teacher rating form (TRF) validation studies determine what may properly be inferred from TRF scores. Such studies provide evidence of the usefulness of TRFs. Studies of TRF generalizability are concerned with determining the conditions under which student ratings are valid. For example, will the TRF be equally useful under condition A as condition B? Studies of TRF generalizability may focus on three sets of variables: student characteristics, teacher characteristics, or situational characteristics. Researchers have attempted to determine if TRF scores are influenced by characteristics of student raters because student characteristics such as age, sex, and ability vary within and between classes. Since teacher and situational characteristics vary between classes it is also important to study TRF generalizability across these sets of factors. Studies of the effects of teacher characteristics on TRF validity may either identify teacher variables which affect TRF scores or teacher characteristics which facilitate student learning, or both. These studies may be useful in identifying effective instructors and improving college teaching. In addition, studying the generalizability of TRF scores across settings may identify situational variables, such as class size or class time, which affect student ratings or student learning, or both. Typically, TRF generalizability studies merely examine the relationship between TRF scores and student, teacher, or situational characteristics. For example, Doyle and Whitely (1974) report data to suggest that student ratings are predominantly unrelated to student characteristics. Unfortunately, such findings may or may not be evidence of generalizability depending on the goals of the TRF researchers.

Deciding on the goals of teacher evaluation answers the question, "What should the TRF measure?" There are many different goals against which a TRF

may be evaluated. For example: (a) Do the TRF items elicit a representative sample of student descriptions of their instructors? (b) Does the TRF predict teacher effectiveness as measured by some criterion? (c) Does the TRF predict teacher-produced differences in students on some criterion measure?

If one were interested in assessing representative student descriptions as the only goal of teacher evaluation it would be important to insure that the TRF items elicit a fair sample of all dimensions of student descriptions about instructors that are of interest to the assessor. Thus, the relationship between student ratings and student learning would not be of interest. If, however, one were concerned with predicting teacher effectiveness as measured by student learning, the correlation between ratings and learning would be critical while the representativeness of the responses elicited by the TRF items would be unimportant. If one were interested in how well the TRF predicts teacher-produced differences in student learning, finding a TRF-learning correlation would be insufficient evidence for the adequacy of the TRF unless differences in student learning could be attributed to teachers.

Choosing among the different goals of teacher evaluation is a philosophical, not a scientific, matter. However, after one has selected the particular goal(s) of teacher evaluation there are definite validation procedures to follow. Each goal requires a different method for establishing TRF validity derived from a unique validity model.

For TRF researchers interested in representative student descriptions as the goal of teacher evaluation it must be shown that the TRF items can elicit a representative sample of the performance domain of interest. For

example, Hildebrand and Wilson (1970) surveyed 338 students for their descriptions of the characteristics of effective teachers. From the student responses a 91 item pool was generated. Factor analyses of the responses of 1015 students to the items about their instructor yielded five dimensions, each represented by several items. The Hildebrand and Wilson TRF is valid to the extent that these procedures resulted in a set of items which can elicit a representative sample of student descriptions about effective teachers. Such a validity model resembles the content validity model as described in the American Psychological Association (1974) test standards manual.

As Leventhal (1975) notes, many studies which investigate the relationship of TRF scores to some validity criterion employ the same core design. Students are non-randomly assigned to multiple sections of the same course. Towards the end of the course students rate their instructor on a TRF and students are measured on a validity criterion, such as a common final exam. Finally, a single TRF mean and a single exam mean are computed for each section, and a correlation coefficient is computed across sections between the TRF means and exam means. The correlation coefficient is then used as evidence concerning TRF validity. This validity model resembles the criterion-related validity model (American Psychological Association, 1974).

A somewhat different validity model requires that TRF scores correlate with teacher characteristics that theoretically affect a validity criterion such as student learning. Unlike the previous model, this model may (depending on the theory) require that TRF ratings predict variation in the validity criterion that is caused by teachers. Such a model resembles the construct validity model (American Psychological Association, 1974).

Many goals of teacher evaluation, other than the three described, are possible. For example, a TRF researcher might be interested in showing that representative student descriptions predict teacher effectiveness as measured by some criterion. Some TRF researchers may be interested in showing that representative student descriptions are not influenced by characteristics presumed to be unrelated to teacher effectiveness. Here, researchers are testing, in part, the validity of the student opinion construct. That is, if it can be shown that student or teacher characteristics (e.g., age, sex, or "personality") or other variables (e.g., expected grade in course), presumed to be unrelated to teacher effectiveness, affect the opinions that students form about teachers then the validity of TRF scores may be questioned. This will occur even if the TRF scores form a representative sample of descriptions about effective teachers.

As already noted, some validity models do not require any evidence that a TRF correlate with a criterion such as student learning for the TRF to be valid. For these models, if student ratings and student learning are found to be uncorrelated, Whitely (1975) correctly suggests that these measures may be used as independent sources of information about teaching. Unfortunately, the use of a TRF which is unrelated to some aspect of student learning would probably unsettle most college instructors. Nevertheless, such validity models are used by some TRF researchers.

Leventhal (1975) notes that the goal of most TRF researchers is not merely to demonstrate that TRF scores correlate with some validity criterion but (due to implicit theories about how effective teachers affect students) to show that TRF scores correlate with <u>teacher-produced</u> differences in the criterion measure. Note that this research goal does not suggest that TRF

items must elicit a representative sample of student descriptions.

For researchers interested in whether ratings predict teacher-produced differences in the criterion measure, a suitable generalizability study would investigate the relationship between student characteristics and the correlation between TRF scores and teacher-produced student learning, or some other criterion. A TRF can only be considered generalizable if the validity coefficient is unaffected by student characteristics. Put another way, generalizing a TRF amounts to repeating a validity study for each student characteristic one wants to generalize to.

Consider the simulated data presented in Table 1. Cases 1 and 2 represent data comparable to Doyle and Whitely's (1974). That is, the student characteristic does not correlate with TRF scores. But the TRF is generalizable in Case 1 only. In Case 1, the correlations, computed for each level of the student characteristic, between mean TRF scores and mean achievement scores are not different. In Case 2, the student characteristic affects student learning in such a way that the correlations between TRF scores and achievement vary. This result is not unlikely with such student characteristics as intellectual ability or motivation to learn. For example, classes of bright students may perform well with little regard to the efforts of their instructor but the classes may also award very different TRF scores depending on certain instructor characteristics. Alternately, classes of dull students may perform poorly with little regard to the efforts of their instructors, yet give out very different teacher ratings.

Doyle and Whitely concluded that TRF scores are unaffected by student characteristics and are therefore generalizable. The simulated data for Cases 1 and 2 suggest that this conclusion will be correct only if the validation

criterion is similarly unaffected. But Doty (1967) demonstrated that various student characteristics correlate with student achievement. Hence the relationship between TRF scores and achievement scores may vary due to student characteristics.

Cases 3 and 4 represent simulated data for a student characteristic that does correlate with TRF scores. Nevertheless, the TRF is generalizable in Case 3 because the correlations between student ratings and achievement scores are unaffected by different levels of the student characteristic. For example, academic level of students was found to correlate with TRF scores (Gage, 1961). Nevertheless, one might suspect a TRF to be generalizable since academic level of student should also correlate with student learning.

When TRF scores are generalizable but student characteristics affect the scores (Case 3), the ratings can be made comparable from teacher to teacher by "correcting" them for the student characteristics. It would be misleading to compare the ratings of two professors whose classes were different with regard to important student characteristics without adjusting the ratings. Once the validity of a TRF has been established for each level of a student characteristic, TRF scores can be corrected for student characteristics in much the same way as IQ scores on intelligence tests are corrected for age of testee.

The preceding discussion has illustrated how to determine whether a TRF which is designed to measure teacher-produced student learning is generalizable across students. In addition to student characteristics, concern is often expressed that student evaluations should not be influenced by the personality characteristics of the instructor. This issue is, in part, a definitional one concerning the exclusion of personality characteristics from the construct

"effective instruction." Researchers who argue that an instructor's personality should not affect the ratings he/she receives imply that personality style is not a dimension of good teaching. But, consider the TRFs based on factor analytic studies (e.g., Isaacson, McKeachie, Milholland, Lin, Hofeller, Baerwaldt and Zinn, 1964) which do include personality dimensions such as "rapport" or "interaction." Here the implicit assumption is made that teacher personality may be an important part of effective instruction.

Studies which attempt to show a relationship between teacher personality and student ratings typically do not include a student achievement measure (e.g., Sherman & Blackburn, 1975); hence, such studies fail to provide evidence of whether TRFs predict teacher-produced student learning. To demonstrate that a TRF does not generalize it must be shown that either: (1) differences in teacher personality cause student ratings to vary and these differences are unrelated to student achievement, or (2) differences in teacher personality cause student achievement to vary and these differences are unrelated to student ratings. In other words, teacher personality must affect the relationship between TRFs and learning for TRF generalizability to be affected.

In addition, researchers may also be interested in TRF generalizability across situational variables (e.g., class time, class size, or different universities). The design and analysis of such studies is similar to that presented for generalizing across student and teacher characteristics. Evidence of whether TRFs predict teacher-produced student learning across situational factors is gained by examining the effects of setting variables on the TRF validity coefficient.

For researchers who are interested in whether TRFs correlate with some

criterion (not necessarily affected by teachers), generalizability studies should be similarily designed. That is, the validity coefficient must be shown to be unaffected by the characteristic one wants to generalize across. However, for researchers whose goal is a TRF which elicits representative student descriptions, a criterion measure is not employed so no validity coefficient is ever computed. For these researchers showing that the representativeness of the responses elicited by TRF scores is unaffected by the characteristics one wants to generalize across, is sufficient evidence of TRF generalizability. For researchers interested in showing that student descriptions are not influenced by characteristics presumed to be unrelated to teacher effectiveness, finding that TRF scores are unaffected by the characteristics one wants to generalize across is sufficient evidence of TRF generalizability. For example, Doyle and Whitely's (1974) research provides good evidence that student descriptions of effective teaching are not affected by certain student characteristics.

In conclusion, for TRF researchers interested in the relationship of TRF scores to teacher-produced student learning, appropriate generalizability studies should include both a TRF and a learning measure. Since in the past, these researchers have failed to study the generalizability of the TRF validity coefficient, the generalizability of TRFs is still an open question.

References

American Psychological Association. Standards for educational and psychological tests. Washington: American Psychological Association, 1974.

Cronbach, L.J. and Meehl, P.E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Doty, B.A. Teaching method effectiveness in relation to certain student characteristics. Journal of Educational Research, 1967, 8, 363-365.

Doyle, K.O., Whitely, S.E. Student ratings as criteria for effective teaching. American Educational Research Journal, 1974, 11, 259-274.

Gage, N.L. The appraisal of college teaching. Journal of Higher Education, 1961, 32, 17-22.

Hildebrand, M. and Wilson, R.C. Effective university teaching and its evaluation. Grant Number OE-6-10-106. Project Number 5-0248-20-6 (University of California, Davis). Washington, D.C.: U.S. Office of Education, 1970.

Isaacson, R.L., McKeachie, W.J., Milholland, J.E., Lin, Y.G., Hofeller, M., Baerwaldt, J.W. and Zinn, K.L. Dimensions of student evaluations of teaching. Journal of Educational Psychology, 1964, 55, 344-351.

Leventhal, L. Teaching rating forms: Critique and reformulation of previous validation designs. Canadian Psychological Review, 1975, 16, 269-276.

Sherman, B.R. and Blackburn, R.T. Personal characteristics and teaching effectiveness of college faculty. Journal of Educational Psychology, 1975, 67, 124-131.

Whitely, S.E. On the invalidity of construct validity for student ratings. Paper presented at the annual meeting of the American Educational Research Association, April, 1975.

TABLE 1

Simulated TRF Generalizability Data[a]

LEVEL OF STUDENT CHARACTERISTIC	TEACHER	STUDENT CHARACTERISTIC DOES NOT AFFECT TRF		STUDENT CHARACTERISTIC AFFECTS TRF	
		MEAN TRF SCORE	MEAN ACHIEVEMENT SCORE	MEAN TRF SCORE	MEAN ACHIEVEMENT SCORE
Hi	A	1.0	1.0	1.1	1.1
	B	3.0	3.0	1.3	1.3
	C	5.0	5.0	1.5	1.5
Med	A	1.0	1.0	3.1	3.1
	B	3.0	3.0	3.3	3.3
	C	5.0	5.0	3.5	3.5
Lo	A	1.0	1.0	5.1	5.1
	B	3.0	3.0	5.3	5.3
	C	5.0	5.0	5.5	5.5
		CASE 1. TRF IS GENERALIZABLE		CASE 3. TRF IS GENERALIZABLE	
Hi	A	1.0	1.1	1.1	1.0
	B	3.0	1.3	1.3	3.0
	C	5.0	1.5	1.5	5.0
Med	A	1.0	3.5	3.5	1.0
	B	3.0	3.3	3.3	3.0
	C	5.0	3.1	3.1	5.0
Lo	A	1.0	5.3	5.3	1.0
	B	3.0	5.1	5.1	3.0
	C	5.0	5.5	5.5	5.0
		CASE 2. TRF IS NOT GENERALIZABLE		CASE 4. TRF IS NOT GENERALIZABLE	

[a]Table 1 shows mean TRF and mean achievement scores for students who are high, medium and low on a student characteristic and who are assigned to different teachers. Scores in the table apply for either of two kinds of student assignment: (1) heterogeneous classes where each teacher (A, B, and C) receives high, medium and low students, or (2) homogeneous classes where each teacher receives high students, then teaches another class of medium students, then another of low students. The choice of homogeneous versus heterogeneous classes for investigating TRF generalizability may depend upon the specific field situation of interest and is beyond the scope of this paper.